

Now that we have understood the foundational concepts, it is time to begin our hands-on practice using **Amazon Bedrock**.

- Before starting, it is important to ensure that the selected region in the AWS console is set to **US East (N. Virginia)** – `us-east-1`, as this region currently provides the best support for Bedrock.
- After verifying the region, the next step is to open **Amazon Bedrock** in the AWS console. Upon opening the service, the **overview page** will be displayed, introducing various available features

The screenshot shows the AWS console interface for the Amazon Bedrock service. The left sidebar contains navigation links for Getting started, Foundation models, Playgrounds, Builder tools, and Safeguards. The main content area displays an 'Introducing Prompt routers' banner, followed by the 'Overview' section which highlights 'Foundation models'. It mentions that Amazon Bedrock supports over 100 foundation models from industry-leading providers and emerging leaders. Below this, there are buttons for 'View Model catalog' and 'Discover marketplace models'. To the right, a 'Model spotlight' box features 'Anthropic's Claude', describing it as a model that can handle various tasks like generating text and images. At the bottom, there are sections for 'Chat / Text' and 'Image / Video', each with a brief description and a 'Discover' button. The footer includes links for CloudShell, Feedback, and various AWS terms like Privacy, Terms, and Cookie preferences.

Amazon Bedrock hosts models from multiple AI providers. As of now, there are **eight providers** available, with more likely to be added in the future. The course focuses on the most important ones.

The screenshot shows the 'Providers' section of the Amazon Bedrock console. On the left, a sidebar lists various services and tools available in Amazon Bedrock. The main area displays a grid of provider cards. Each card includes the provider's logo, name, and model count. The cards shown are:

- AI21 labs**: 3 models | Serverless
- Amazon**: 15 models | Serverless
- Anthropic**: 10 models | Serverless
- (partially visible): A green and orange blob-like logo, a blue whale logo, and a blue infinity symbol.

When a provider, such as **Amazon**, is selected, a dedicated page is displayed containing:

The screenshot shows the detailed provider page for **Amazon**. The sidebar on the left remains the same as the previous page. The main content area is divided into sections:

- Provider details**: Includes a small icon of a server or database and a text block about Amazon's history in AI and ML.
- Spotlight**: Features a callout for **Introducing Amazon Nova**, describing it as a new generation of SOTA foundation models. It includes a "Request model access" button.
- Models overview**: Provides an overview of Amazon Nova models available on Amazon Bedrock, listing five models: Micro, Lite, Pro, Canvas, and Reel.

- A **spotlight section** and an overview of the provider
- A **description of available models**, such as:
 - *Nova Micro*: A **text-only model**
 - *Nova Pro*: A model with a broader category range

The screenshot shows the AWS Amazon Bedrock console. On the left, a sidebar lists various sections: Getting started, Foundation models, Playgrounds, Builder tools, and Safeguards. The 'Models overview' section is highlighted with a red border. It contains a list of Amazon Nova models: Nova Micro, Nova Lite, Nova Canvas, Nova Reel, Nova Sonic, and Nova Pro. A callout box highlights 'Amazon Nova models on Amazon Bedrock are:' followed by a bulleted list: Amazon Nova Micro, Amazon Nova Lite, Amazon Nova Pro, Amazon Nova Canvas, and Amazon Nova Reel.

At the bottom, if you scroll down, you have access to all the models:

The screenshot shows the AWS Amazon Bedrock console with the 'Models (15)' section visible. The sidebar remains the same. The main area displays 15 models in a grid format. The models listed are: Nova Sonic (New), Nova Canvas, Nova Lite, Nova Micro, Nova Pro, Nova Reel, Titan Text G1 - Premier, Titan Image Generator..., and Titan Text Embeddings... Each model card includes its name, a brief description, and an 'Available to request' status indicator.

For each model, if you click on **Nova Pro**. You get description about the model itself:

1. Categories

2. Version

Amazon Bedrock > Model catalog > Nova Pro



Nova Pro

By: Amazon | ⓘ Available to request

Nova Pro is a multimodal understanding foundation model. It is multilingual and can reason over text, images and videos.

[Open in playground](#)



Overview

Nova Pro is a multimodal understanding foundation model. It is multilingual and can reason over text, images and videos.

On this page

[Overview](#)
[Usage](#)

Details

Sold by	Amazon
Categories	Agents Chat optimized Code generation Complex reasoning analysis Conversation Math Multilingual support Question answering RAG Text generation Text summarization Translation Text

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

When you scroll down below, you will get details about **Usage**:

AWS | Services | Search [Option+S] | [United States \(N. Virginia\)](#) | [aws-courses](#)

Amazon Bedrock <

- Getting started
 - Overview
 - Providers
- Foundation models
 - Model catalog** New
 - Marketplace deployments New
 - Custom models (fine-tuning, dist...)
 - Imported models
 - Prompt Routers Preview
- Playgrounds
 - Chat / Text
 - Image / Video
- Builder tools
 - Agents
 - Flows
 - Knowledge Bases
 - Prompt Management
- Safeguards
 - Guardrails
 - Watermark detection

Usage

API request

```
{
  "modelId": "amazon.nova-pro-v1:0",
  "contentType": "application/json",
  "accept": "application/json",
  "body": {
    "inferenceConfig": {
      "max_new_tokens": 1000
    },
    "messages": [
      {
        "role": "user",
        "content": [
          {
            "text": "this is where you place your input text"
          }
        ]
      }
    ]
  }
}
```

[Copy code](#)

On this page

[Overview](#)
[Usage](#)

[CloudShell](#) [Feedback](#)

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

So this is some information around API. In this tutorial, we are going to be using Bedrock in the console.

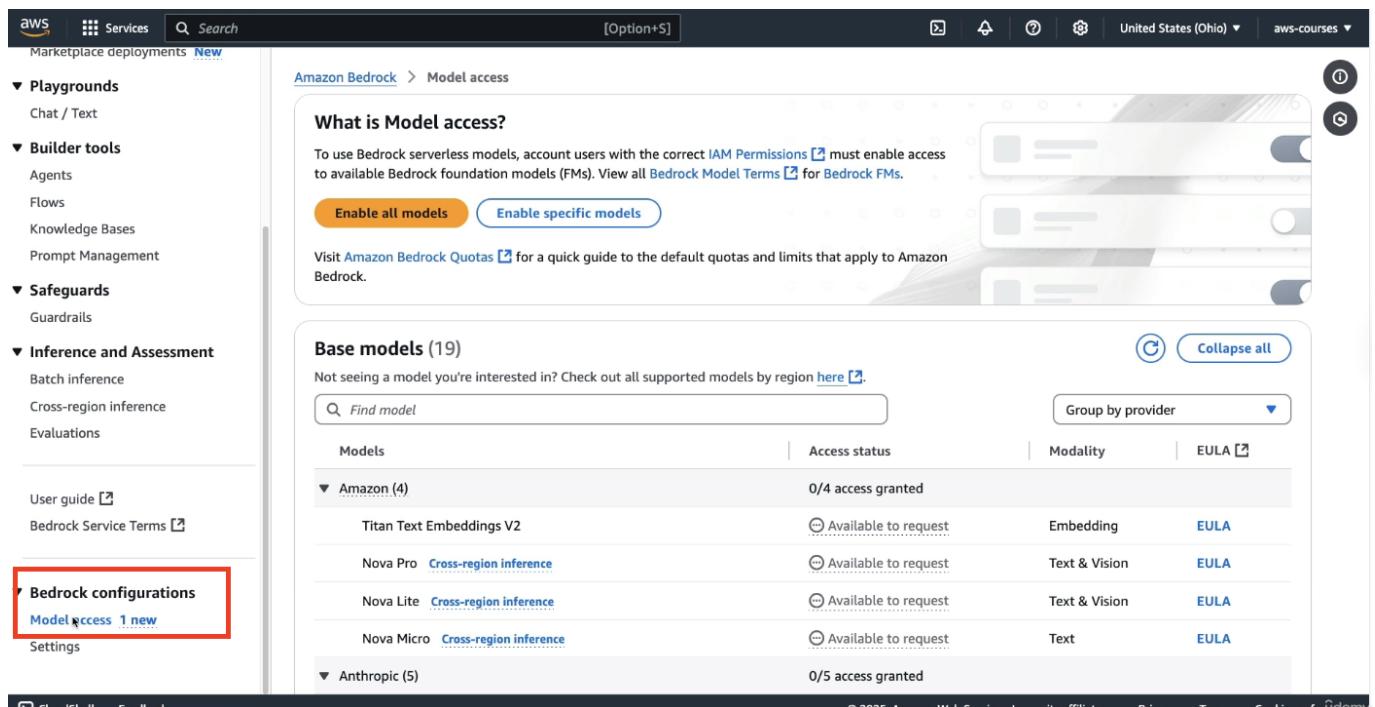
Note that if you want to integrate Bedrock into your Application, you would need to use API request. So the above image about **Usage**, something which AWS tells you how to use API requests for the specific type of model

All of this helpful but **first we need to get access to these models**

Enabling Access to Foundation Models

To interact with any of the foundation models:

- The user must scroll to the bottom left of the Bedrock page and locate the **Model Access** section.
- A full list of models appears, varying slightly between accounts.



The screenshot shows the 'Amazon Bedrock > Model access' page. On the left, there's a sidebar with sections like 'Playgrounds', 'Builder tools', 'Safeguards', 'Inference and Assessment', and 'Bedrock configurations'. The 'Bedrock configurations' section has a red box around it, and the 'Model access' link is highlighted. The main content area has a heading 'What is Model access?' with a note about IAM permissions and a toggle switch. Below it is a section titled 'Base models (19)' with a table. The table has columns for 'Models', 'Access status', 'Modality', and 'EULA'. It lists four models under 'Amazon': 'Titan Text Embeddings V2' (Available to request, Embedding, EULA), 'Nova Pro' (Available to request, Text & Vision, EULA), 'Nova Lite' (Available to request, Text & Vision, EULA), and 'Nova Micro' (Available to request, Text, EULA). There are also sections for 'Anthropic' (5) and 'Meta' (1).

To proceed:

- The “**Enable all models**” option should be selected. This action does **not** incur charges.
- A few clicks are required: select **Next**, then **Submit** to confirm access.

⚠️ Although enabling models is free, **actual usage of models may result in charges**. This course attempts to minimize costs, but users should be aware that **no AI model usage is entirely free**.

Some models grant access **instantaneously**, while others may show as **in progress** and require time for approval. Certain models may prompt for **license agreement** confirmation.

Temporary errors or delays should not be a concern, as enough models will be enabled to complete the hands-on work.

The screenshot shows the 'Edit model access' page in the Amazon Bedrock console. The left sidebar indicates 'Step 1: Edit model access' is selected. The main area displays a table titled 'Base models (19/19)' with columns for 'Models', 'Access status', 'Modality', and 'EULA'. The table lists various models under the 'Amazon' provider, including Titan Text Embeddings V2, Nova Pro, Nova Lite, Nova Micro, Claude 3.7 Sonnet, Claude 3.5 Haiku, Claude 3.5 Sonnet v2, and Claude 3.5 Sonnet. Most models have an 'Available to request' status, while some like Titan Text Embeddings V2 and Nova Pro have '0/4 access granted'. The 'Modality' column shows 'Embedding', 'Text & Vision', 'Text & Vision', 'Text', 'Text & Vision', 'Text', 'Text & Vision', and 'Text & Vision'. The 'EULA' column shows 'EULA' repeated for most models. A search bar at the top allows finding specific models. The bottom of the page includes standard AWS navigation links like CloudShell, Feedback, and a footer with copyright information.

Edit model access

Base models (19/19)

Not seeing a model you're interested in? Check out all supported models by region [here](#).

Find model

Group by provider

Models	Access status	Modality	EULA
Amazon (4)	0/4 access granted		
Titan Text Embeddings V2	Available to request	Embedding	EULA
Nova Pro Cross-region inference	Available to request	Text & Vision	EULA
Nova Lite Cross-region inference	Available to request	Text & Vision	EULA
Nova Micro Cross-region inference	Available to request	Text	EULA
Anthropic (5)	0/5 access granted		
Claude 3.7 Sonnet Cross-region inference	Available to request	Text & Vision	EULA
Claude 3.5 Haiku Cross-region inference	Available to request	Text	EULA
Claude 3.5 Sonnet v2 Cross-region inference	Available to request	Text & Vision	EULA
Claude 3.5 Sonnet Cross-region inference	Available to request	Text & Vision	EULA

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Model access modifications (19)

Models	Modifications
Llama 3.1 405B Instruct	Request access
Llama 3.1 70B Instruct	Request access
Llama 3.1 8B Instruct	Request access
Nova Pro	Request access
Nova Lite	Request access
Nova Micro	Request access
DeepSeek-R1	Request access
Pixtral Large (25.02)	Request access

Terms
By selecting Submit, you are requesting access to the selected third party models through the AWS Marketplace. By doing so, you agree to the seller's pricing terms and End User License Agreements (EULA), and the [Bedrock Service Terms](#). You also agree and acknowledge that AWS may share information about this transaction with the respective sellers, in accordance with the [AWS Privacy Notice](#).

AWS will issue invoices and collect payments from you on behalf of the seller through your AWS account. Your use of AWS services is subject to the [AWS Customer Agreement](#) or other agreements with AWS governing your use of such services.

Cancel Previous Submit

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Using the Chat Playground for Text Models

Once model access is enabled:

- The user should navigate to the left sidebar and open the **Playgrounds** section.

Amazon Bedrock

- Getting started
 - Overview
 - Providers
- Foundation models
 - Model catalog [New](#)
 - Marketplace deployments [New](#)
 - Custom models (fine-tuning, dist...)
 - Imported models
 - Prompt Routers [Preview](#)
- Playgrounds
 - Chat / Text**
 - Image / Video
- Builder tools
 - Agents
 - Flows
 - Knowledge Bases
 - Prompt Management
- Safeguards
 - Guardrails
 - Watermark detection

Amazon Bedrock > Model access

What is Model access?

To use Bedrock serverless models, account users with the correct IAM Permissions must enable access to available Bedrock foundation models (FMs). View all [Bedrock Model Terms](#) for Bedrock FMs.

[Modify model access](#)

Visit [Amazon Bedrock Quotas](#) for a quick guide to the default quotas and limits that apply to Amazon Bedrock.

Base models (49)

Not seeing a model you're interested in? Check out all supported models by region [here](#).

Models	Access status	Modality	EULA
▼ AI21 Labs (3)	3/3 access granted		
Jamba 1.5 Mini	Access granted	Text	EULA
Jamba 1.5 Large	Access granted	Text	EULA
Jamba-Instruct	Access granted	Text	EULA
▼ Amazon (14)	14/14 access granted		
Titan Embeddings G1 - Text	Access granted	Embedding	EULA

CloudShell Feedback [Feedback](#) © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

- Inside Playgrounds, selecting **Chat (text)** opens the interface for prompt-based model interaction.

For this exercise:

- The **Single Prompt** option should be selected (it means you can only send 1 query and there won't be any follow ups on that prompt)

Amazon Bedrock > Chat / Text playground

Mode **Single prompt**

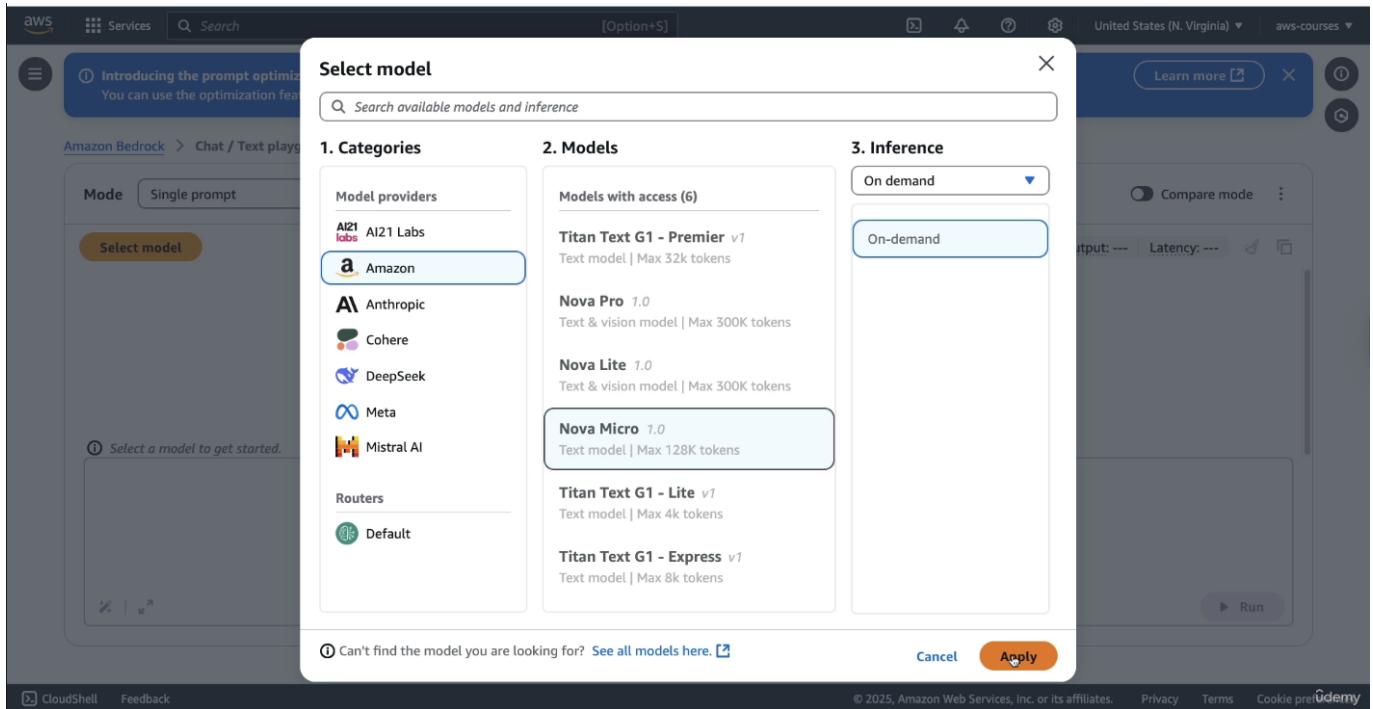
[Select model](#)

Select a model to get started.

Input: --- Output: --- Latency: ---

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

- You need to click on **Select model**, and click on **Amazon Nova Micro**. It is the **cheapest and most accessible on-demand model** from AWS.
- Once the model is applied, the user can proceed to input a prompt.



To test the model, a basic prompt is used:

!"What is AWS?"

This prompt is sent through Amazon Bedrock to the **Nova Micro model**, which processes it and returns a structured answer. The model generates a list of key features related to AWS in a numbered format.

Along with the answer, Bedrock displays several **important metadata values**:

- The **number of input tokens** used
- The **number of output tokens** generated
- The **latency**, which refers to the time it took for the model to generate the response

These statistics are important for understanding **pricing and performance**.

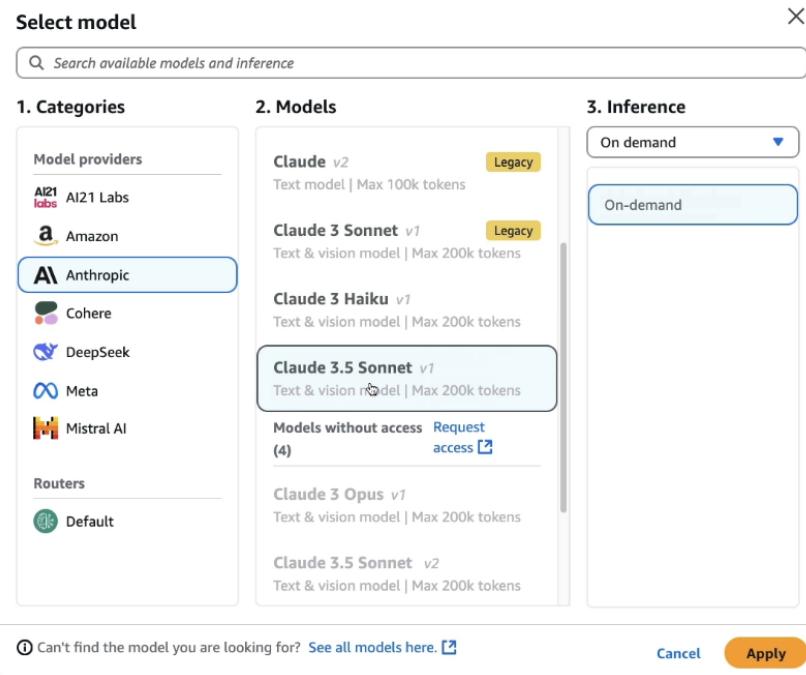
The screenshot shows the Amazon Bedrock interface with the 'Mode' dropdown set to 'Single prompt'. On the left, there's a 'Configurations' section with a 'Nova Micro 1.0' model selected. Below it is a 'System prompts' section with a placeholder 'Add system prompts'. Under 'Randomness and diversity', two sliders are shown: 'Temperature' at 0.7 and 'Top P' at 0.9. On the right, a large text area displays a response from the model. A red box highlights the 'Input tokens' section, which shows 'Input: 4', 'Output: 510', and 'Latency: 2368 ms'. The response text discusses AWS's global access, developer tools, cost management, and popularity.

In addition to single-prompt interactions, Amazon Bedrock offers a **chat mode**, which allows users to conduct ongoing conversations with the model while preserving context.

This screenshot shows the same interface but with the 'Mode' dropdown set to 'Chat'. The 'Chat' option is highlighted with a red box. The rest of the interface and the response text are identical to the previous screenshot.

In this mode:

- The user can still select models such as **Amazon Nova Pro** or **Claude 3.5** from Anthropic model (**Yes!!! You need to again select the model**)



- Once a model is selected, a prompt such as “*What is AWS?*” can be entered again to observe how the response changes. Upon clicking **Run**, a response is generated by the selected model.

It is important to note that:

- Responses will vary across different models and even across runs on the same model. (As seen in Amazon Bedrock Overview Lecture that LLMs are **Non-deterministic**)
- Writing style and quality differ significantly between models.
- Less expensive, lower-performance models tend to produce answers with slightly reduced quality compared to high-performance, premium models.
- In the chat mode, we have more parameters as you can see on RHS

The screenshot shows the Amazon Bedrock interface in Chat mode. At the top, there's a blue banner with the text "Introducing prompt caching" and a note about it automatically storing user inputs and model responses. Below the banner, the "Mode" dropdown is set to "Chat". On the left, there are several configuration sliders: "Top P" (set to 0.999), "Top K" (set to 250), and "Length" (with "Maximum length" set to 2000). A "Stop sequences" input field is also present. In the center, a text input area contains the placeholder "Write a prompt and choose Run to generate a response." Below the input area is a "Run" button. At the bottom of the interface, there are links for CloudShell, Feedback, and various AWS terms like Privacy, Terms, and Cookie preferences.

- Also in chat mode, you will find attachments, not every model will provide this option.

For each chat interaction, Bedrock provides:

- The number of **input tokens**
- The number of **output tokens**
- The **latency**, which indicates the time taken for the response to be generated

Generating Images Using Amazon Bedrock

Amazon Bedrock also supports **image generation**, enabling the creation and modification of visual content directly through its interface.

Introducing prompt caching
Prompt caching automatically stores user inputs and model responses, maintaining context for faster retrieval from the model and reduced costs.

Learn more

Amazon Bedrock > Chat / Text playground

Mode: Chat | Compare mode

Configurations: Claude 3.5 Son... v1

Input: 12 | Output: 361 | Latency: 10289 ms

增加或减少资源基于需求。

- 全球基础设施: AWS 在世界各地拥有数据中心(称为区域), 提供低延迟访问和数据冗余。
- 安全: AWS 提供 robust 安全功能并符合许多行业标准。
- 通用性: 它服务于广泛的行业和用例, 从小型初创公司到大型企业。
- 创新: AWS 定期推出新服务和更新现有服务以跟上技术进步。
- 流行的服务: 一些著名的 AWS 服务包括 EC2 (虚拟服务器), S3 (存储), RDS (托管数据库), 和 Lambda (无服务器计算)。

Write a prompt. Press Shift + Enter to add a new line. Press Enter to generate a response.

Run

For this demonstration:

- The **Nova Canvas** model is selected and applied for image generation.

Select model

Search available models and inference

1. Categories

Model providers

- Amazon**
- Stability AI

2. Models

Models with access (5)

- Titan Image Generator G1 v1**
Image model | Max 512 characters
- Titan Image Generator G1 v2**
Image model | Max 512 characters
- Nova Canvas 1.0**
Image model | Max 1024 characters
- Nova Reel v1**
Video model | Max 512 characters
- Nova Reel v1.1**
Video model | Max 512 characters

3. Inference

Select model to show inference options.

Can't find the model you are looking for? See all models here.

Cancel Apply

- Several features are available (under **Action** tab) beyond simple image creation, including:
 - Generating variations of an image
 - Removing or replacing objects

- Changing or removing backgrounds

A sample prompt is provided to test the image generation capabilities:

“Generate an orange backpack with the AWS logo on it on someone that is teaching a live AWS workshop.”

The model processes the prompt and returns an image:

The screenshot shows the Amazon Bedrock Image generation playground interface. On the left, there's a sidebar with 'Nova Canvas 1.0' selected. The main area has three generated images: a person from behind wearing an orange backpack, a person from the side wearing an orange backpack, and a close-up of an orange backpack with the AWS logo.

This example shows that while image generation is functional and advanced, not all prompts will be fulfilled perfectly, especially when they involve **multiple layers of semantic detail**.